DOCUMENT RESUME

ED 149 538 EC 103 854

AUTHOR Alvir, Howard P.
TITLE What Criteria Should Be Used to Identify Successful Teaching in Special Education and in Occupational Education of the Handicapped.
PUB DATE 77
NOTE 15p.
AVAILABLE FROM Claire Gelinas, 27 Norwood Street, Albany, New York 12203 (send stamped self addressed envelope for information)

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Evaluation Methods; Exceptional Child Research; *Handicapped Children; *Research Methodology; *Statistical Analysis

ABSTRACT

Offered is a method of statistically analyzing questions concerned with identifying criteria used in evaluating teaching methods used with the handicapped. The method is demonstrated with the question of the effect of teacher presenters knowing beforehand the criteria on which their presentations will be judged. The evaluation design is described in terms of variables (four evaluation dimensions), evaluators, matched pairs, the research question, and the correlation process. A glossary of statistical terminology is used. Advantages of the method are seen to include a built-in control group and experimental group design and low cost as calculations can be performed on a hand held calculator. Results of applying the method to the demonstration question indicate that prenotification of evaluation criteria usually increases teaching effectiveness. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED149538

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Title

What Criteria Should Be Used to Identify Successful Teaching in Special Education and in Occupational Education of the Handicapped

© Copyright 1977
Howard F Alvir

27 Norwood St
Albany, N. Y. 12203

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Howard Alvir

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

EC 103852

PREFACE

WHAT THIS DOCUMENT IS:

This document is a description of the literature searching methodology, design, data analysis, and application of a simplified and practical approach to answering the following question:

> What criteria should be used to identify successful teaching in special education and in occupational education of the handicapped?

The LITERATURE SEARCH identifies four different types of criteria that have been unable to answer the following question conclusively:

> Will teachers do a better job with the handicapped if they know beforehand the criteria on which their teaching will be evaluated?

The METHODOLOGY of this evaluation design answers the above question scientifically by rephrasing it into the obserable research question:

> Can uncomplicated one page evaluation forms identify which teacher in a matched pair of educators was prenotified about the evaluation criteria?

The DESIGN of this evaluation was tested with 10 teachers, 100 evaluators, and 200 one page forms. It rests on the following assumptions for replication:

A. Matched pairs of teachers can be found.
B. Even the best matched pairs of teachers will contain some unequal pairs when significant variables are measured with pretests and posttests.
C. All unmatched pairs (in step B above) must be eliminated from the data analysis as inconclusive and contaminated data.

This design led to a very simple approach to DATA ANALYSIS that possessed a built-in CONTROL GROUP and EXPERIMENTAL GROUP DESIGN. Here it is in a few highlights:

A. CONTROL GROUP: Some teachers received no prenotification.
B. INTERNAL CONTROL GROUP: Some teachers received prenotification of only one evaluation criterion.
C. EXTERNAL CONTROL GROUP: Only two evaluators knew <u>who</u> had received prenotifcation and <u>what</u> had been prenotified in each case. The statisticians had to rely strictly on the evaluation forms to identify <u>who</u> had <u>what</u> prenotification.
D. EXPERIMENTAL GROUP: Some teachers were in <u>both</u> an experimental group (which had been prenotified about one of the four evaluation forms) <u>and</u> a control group (which received no prenotification about three of the four evaluation forms).

The APPLICATION of this research led to the following conclusions:

A. On most criteria, prenotification is recognizable.
B. On some criteria, prenotification is not recognizable.
C. The criteria where prenotification is effective outnumber the other criteria.

NOTE: For several reasons, the target population of this research and the criteria where prenotification is effective have not been identified:

A. The purpose of this document is methodological, not etiological.
B. This methodology is currently being replicated in several sites with a larger population (N=200 teachers).
C. It is hoped that interested special educators will replicate this methodology and share their findings.
D. It is felt that the small population (N=10 teachers) was adequate to test out a methodology but inadequate to test out etiological (cause-and-effect) speculations on where prenotification is associated with effectiveness.

Contents

ABSTRACT ---------- 1

INTRODUCTION ---------- 2

Either/or Previous Research

Which Type Previous Research

Which One Previous Research

Which Criteria Previous Research

Matched Pair PRENCTIFICATION Research

Evaluation Design ---------- 4

Variables

Evaluators

Matched Pairs

Research Question

Correlation

What We Did and How We Did It ---------- 5

Glossary (For Replication) ---------- 6

Sample Data No. 1 (Conclusive Results) ---------- 7

Sample Data No. 2 (Inconclusive Results) ---------- 8

How You Can Use This Data ---------- 9

Application

Conclusion ---------- 10

ABSTRACT

While working with handicapped students, one question recurred time and time again: "What criteria should be used to identify successful teaching?"

We answered this question with the help of a panel of experts from the following areas:

A. Subject matter (e.g. occupational education)
B. Special education (e.g. counselors, teachers, administrators, supervirors, professors)
C. Curriculum (e.g. writers, editors, experts, supervisors)
D. Instruction (e.g. teachers, supervisors, professors, principals, superintendents)
E. Employment (e.g. counselors, employees, employers, agencies)

Without announcing the criteria, we compared method R with S and so on. We found little significant difference.

Next, we announced our criteria to half of the teachers in the research. Then, came the surprise. Most of the prenotified teachers did a better job as measured against the criteria on which they had received prenotification.

We wanted to QUANTIFY our results. Yet nowhere in any stat book could we find an exact fit.

Basically we had:

MATCHED PAIRS,
CONTROL GROUPS, and
EXPERIMENTAL GROUPS.

We wanted to find out statistically, "Which twin had the TONI (that is, the prenotification)?"

The following pages SUMMARIZE what we did and how we did it. They are presented here to save you time and money the next time you face a similar evaluation problem.

There's a secret we've saved until now. All of the math (except the interpretation of Data No. 1 and 2) were done with a pre-programmed calculator that cost less than $50. When we got done, we found out that other brands could have done the job for less than $40. Maybe, this will encourage you to give it a try.

INTRODUCTION

It's EASY to ask, "Will presenters do a better job if they know beforehand the criteria on which their presentations will be judged?"

It's IMPORTANT to ask this question. If the answer is YES, then PRENOTIFICATION is a proven cost effective way to increase performance. If the answer is NO, then cancelling PRENOTIFICATION is a painless way to save time and trouble.

A survey of evaluation research reveals that this question has been asked in several different ways. Unfortunately, each phrasing of the question has still left plenty of room for error to creep into even the most scientific answer.

Either/Or Question (Previous Research Type 1)

One way to ask the question is to inquire, "Does TOTAL prenotification produce better evaluation results than ZERO prenotification?"

Which Type Question (Previous Research Type 2)

Another way is to inquire, "Which type of prenotification produces the best evaluation results?"

Which One Question (Previous Research Type 3)

Still another way is to inquire, "Which type of person profits most during evaluation as a consequence of prenotification?"

Which Criteria Question (Previous Research Type 4)

Another variant asks, "On which criteria does prenotification produce the most measurable results?"

Matched Pair Question (This Cumulative Research Type 5)

Putting all these questions together for the first time leads this researcher to inquire, "Can uncomplicated evaluation forms spot which TWIN in a matched pair has PRENOTIFICATION?"

As in Type 1 Prenotification designs, the experimental group, that is, the X-group, receives TOTAL prenotification. Unlike Type 1 Prenotification, the prenotification is limited to only one criterian among several criteria actually used to evaluate. No one in the control group, that is, the Y-group, receives any prenotification.

As in Type 2 Prenotification, attention is given to the type of prenotification and to the manner in which prenotification takes place. Unlike Type 2 Prenotification, the prenotification is made exactly the same for all participating presenters in the X-group. All members of the Y-group are treated alike, that is, no member of the Y-group gets any kind of prenotification.

As in Type 3 Prenotification, attention is given to personal traits of members of both the X-group and the Y-group. Unlike Type 3 Prenotification, the X-group and the Y-group were compared on a matched pair basis.

As in Type 4 Prenotification, attention is given to a variety of criteria. Unlike Type 4 Prenotification, each criterion is isolated one at a time for an analysis of the effects of prenotification upon it as measured in the X-group evaluations counter-balanced by the Y-group evaluations.

The following EVALUATION DESIGN develops a rationale for asking SCIENTIFICALLY, "Which TWIN has the PRENOTIFICATION?"

Introduction to the Evaluation Design

Four evaluation dimensions including items, questions, and criteria were identified and coded with A, B, C, and D. Each criterion was measured on a separate form coded A, B, C, or D. Each form had 10 evaluation items. A score of between 0 and 10 was possible on each form.

One hundred evaluators evaluated 10 matched pairs of presenters. This produced 200 evaluation forms almost equally distributed among the four criteria except that all evaluators evaluated one criterion for the purpose of determining inter-rater reliability.

Ten matched pairs of presenters were evaluated to pinpoint the differential effects of prenotification. All members of the matched pairs were equated on criteria A, B, C, and D except that only one member of each pair received prenotification of one of the exact criteria to be used in this evaluation.

With 100 evaluators submitting 200 forms, it was possible to have each presenter evaluated by several evaluators and to have each presenter evaluated on each criterion A, B, C, and D. The experimental, or X-group, received prenotification on one criterion; 100 forms came from the X-group. The control, or Y-group, received no prenotification; 100 forms came from the Y-group.

The following outline will help structure the logistics of this design for possible replication.

Evaluation Design

Variables

Four evaluation dimensions were identified. Each dimension has its own evaluation form identified below:

Dimension	Form
Overall	A
Specifics	B
Theory	C
Practice	D

Evaluators

One hundred evaluators evaluated 10 matched pairs of presenters with the following distribution:

	Number of Evaluators	Number of Forms Used			
		A	B	C	D
	34	34	34		
	33	33		33	
	33	33			33
Subtotal:	100	100	34	33	33
Total:	100 Evaluators	200 Forms			

Matched Pairs

Each presenter in each matched pair was observed by different evaluators in such a way that Forms A, B, C, and D were used to evaluate each presenter.

One of the presenters in each matched pair was shown one of the evaluation forms before the evaluation. This defined the experimental or X-group.

The other presenter in each matched pair was not shown any of the evaluation forms beforehand. This defined the control or Y-group.

The Research Question

This X-group and Y-group design was developed to answer the question:

> "What are the measurable effects
> of showing evaluation forms to
> presenters before their presentations?"

Correlation

Each item on 100 forms from the X-group was correlated with 100 forms form the Y-group.

Correlations, means, standard deviations, differences between means, and standard deviations of the differences were calculated for the following combinations of variables:

A for groups X and Y
B for groups X and Y
C for groups X and Y
D for groups X and Y

What We Did and How We Did It

We took the evaluation forms we received and divided them up into the X-group that received prenotification and the Y-group that received no prenotification. For each group, we computed an average (e.g., an X-mean and a Y-mean), a standard deviation (X-SD and Y-SD), a correlation coefficient (XY-r), the difference between the X-average and the y-average (XY-difference), the standard deviation for this difference (XY-SD), and the significance of the difference (t-test). All of this was done on a pre-programmed hand-held calculator that cost less than fifty dollars. All we had to do was enter the RAW DATA and push the right button for the above data.

For those who want to use a pencil, an adding machine, or a slide-rule, the glossary below provides both definitions and formulas.

After the glossary, sample data no. 1 and 2 give examples of how to display and interpret the data output from an inexpensive calculator. Sample data no. 1 is an example of conclusive data that tells which TWIN had prenotification. Sample data no. 2 is an example of inconclusive data that suggests the allegedly indentical twins are not really matched very well.

Why We Did It This Way

We were primarily interested in determining the criteria necessary to identify successful teaching. Our previous experience confirmed the hunch that "There are many DIFFERENT ways to provide successful teaching. Once the SCORING or evaluation criteria for successful teaching are made public and explicit, each teacher or presenter will choose an approach that can meet these EXPECTATIONS."

We did the evaluation design this way to see if we could indeed experimentally identify which twin in each matched pair had received PRENOTIFICATION of the EVALUATION CRITERIA, that is, of the expectations we had for successful teaching.

GLOSSARY

Term or Abbreviation	Meaning
X-mean	The average of the X-scores expressed in rating units on a scale where 15 is the maximum and 1 is the minimum rating.
X-SD	The standard deviation for the X-scores.
XY-r	The correlation (r) between the X-scores and the Y-scores.
XY-difference	The difference between X-mean and Y-mean.
XY-SD	The standard deviation of the difference: $XY\text{-}SD = \sqrt{(X\text{-}SD)^2 + (Y\text{-}SD)^2}$
$t = \frac{XY\text{-}diff}{XY\text{-}SD}$	A t-test based upon the maximum XY-SD.
$XY\text{-}SE = \frac{XY\text{-}SD}{\sqrt{N}}$	The standard error of the XY-comparisons. Using XY-SE instead of XY-SD in the t-test will give higher levels of significance.

Why This Glossary Is Provided

This glossary is provided to help the reader replicate this evaluation design without wasting time in digesting statistical explanations found elsewhere, without wasting money on computer time which can cost $50 an hour, and without wasting the easily obtained information available from hand-held calculators that cost less than $50 each. Since each model of an inexpensive calculator may have a different word or symbol on its keys, the glossary will help avoid errors in confusing standard deviation (S.D.) keys with significant digit (S.D.) keys. Similarly, the t-test formula as defined in the glossary should not be confused with the t-memory key.

Sample Data No. 1

	X-mean	X-SD	Y-mean	Y-SD	XY-r	XY-diff	XY-SD	t-diff / SD
A.	10	1.0	9	1.1	.88	+1	1.49	.67
B.	8	0.9	7	1.3	.90	+1	1.58	.63
C.	6.5	1.2	6	1.1	.87	+5	1.63	.31
D.*	9	1.3	5	0.9	.63	+4	1.58	2.53
Average	8.4	1.0	8.0	2.3	.82	1.4	1.28	1.09

Sample Interpretation:

1. Variable: In the above data, the asterisk (*) means that the X-group members were shown Form D beforehand. The asterisk localizes the prenotification, whose effects are under study.

2. Reliability: The fact that X-SD, Y-SD, and XY-SD are nearly identical can be interpreted to mean that inter-observer agreement is high for all four variables.

3. Validity: The fact that t-tests for independent variables are insignificant and relatively uniform, can be interpreted to mean that the matched pairs are indeed matched on the independent variables. This adds meaning to any measured differences found in dependent variables.

4. Variance: The fact that the correlation r is lower for the D-variable can be interpreted to mean that the matched pairs differ on the D-variable.

5. Significance: The fact that the t-test is higher for the D-variable can be interpreted to mean that the D-difference is significant.

6. Dead-end: The fact that the average correlation is .82 and that the average t-test is 1.09 can be interpreted to mean that the AVERAGES cause any significant differences to regress to the mean. For this reason, the AVERAGE will not be analyzed.

How to Imitate the Above Sample Data Analysis

The above sample data shows you how to arrange your local data once it has been processed by a calculator or by statistical analysis. Here's what you do: First, fill in all your data except the averages; you don't need the averages which are often misleading. Second, interpret your data as done under the headings of variable, reliability, validity, variance, and significance.

-7-

Sample Data No. 2

	X-mean	X-SD	Y-mean	Y-SD	XY-r	XY-diff	XY-SD	t=diff÷SD
A.	10	.91	6	1.10	.54	4	1.43	2.8
B.	9	1.01	8	.89	.87	1	1.35	0.74
C.*	11	.89	6	.95	.49	5	1.30	3.85
D.	9	.99	8.5	1.2	.93	.5	1.56	.32

Sample Interpretation:

1. Variable: The asterisk(*) means that the X-group members were shown Form C beforehand.

2. Reliability: The fact that X-SD, Y-SD, and XY-SD are nearly identical can be interpreted to mean that inter-observer agreement is high for all four variables.

3. Validity: The fact that t-tests for both B and D are insignificant and nearly identical and that correlations for A and C are significant and nearly identical requires investigation.

4. Variance: The fact that the correlation r values are lower for the A-variable and the C-variable can be interpreted to mean that the matched pairs differ on both variables.

5. Significance: The facts reveal that variables A and C:
   (a) Have high t-test values, i.e., 2.80 and 3.85 which need to be explained.
   (b) Have lower correlations (i.e., 0.54 and 0.49) than the B and D variables.

   These facts can be interpreted to suggest the following hypotheses:
   (a) The alleged matched pairs are not matched on independent variable A.
   (b) The dependent variable C is highly correlated with independent variable A.
   (c) The significant t-test for variable C is due to the fact that the alleged matched pairs are not matched on independent variable A.

6. Caution: The counts in Sample Data No. 1 and No. 2 are based upon different sets of matched pairs in the same population. The pairs in Sample Data No. 1 have been established to be matched on independent variables A, B, and C. Therefore, prenotification conclusions can be justified when D is the dependent variable.

   The pairs in Sample Data No. 2 have been established to be matched on independent variables B and D, and to be mismatched on independent variable A. Therefore, the influence on A on C rather than prenotification is a more likely explanation of the significant t-test in Sample Data No. 2.

How You Can Use This Data

To use this process, here is all you have to do.

First, take your current evaluation forms and count how many evaluation items you have. In both Sample Data No. 1 and 2, only four evaluation items, questions, or criteria were analyzed. They were A, B, C, and D.

Second, decide which evaluation criteria will be used for prenotification, that is, explained beforehand to selected individuals being evaluated. In sample Data No. 1, criterion D was used for prenotification; in Sample Data No. 2, criterion C was used for prenotification.

Third, separate the completed evaluation forms into the X-group with prenotification and into the Y-group without prenotification. In Sample Data No. 1 and 2, each column is clearly labeled as either X or Y.

Fourth, enter your data into an appropriate calculator or compute the mean, standard deviation, and correlation coefficient any way you prefer. You might want to photocopy Sample Data No. 1 or 2 and write in your own calculations alongside either model. Calculate X and Y separately.

Fifth, subtract Y-mean from X-mean to fill in the XY-difference column. For example, 10 minus 9 equals 1.

Sixth, square X-SD and add it to the square of Y-SD and use the square root of this sum to fill in the XY-SD column. For example, 1^2 and 1.1^2 equals 2.21 when rounded; then, the square root of 2.21 equals 1.49 when rounded.

Seventh, divide XY-diff by XY-SD. For example, 1 divided by 1.47 equals 0.67 when rounded.

Eighth, decide whether your results fit the conclusive pattern found in Sample No. 1 or the inconclusive pattern found in Sample Data No. 2.

APPLICATION

When this is done, you will be able to measure how much or how little good was done in your evaluation by prenotification. With this information, you can decide for yourself whether or not prenotification was worth the extra time, cost, and effort. When you've done this, you've set up your own decision oriented evaluation system. You then will have a tool for the rational allocation of scarce resources in your budget.

Conclusion

This research asked the question, "What criteria should be used to identify successful teaching of the handicapped?"

No answer was given. Instead, a method was demonstrated that shows educators how to get this answer on their own. Obviously, this method stresses PROCESS and LOCAL DATA.

The PROCESS is D.O.E.S., that is, a Decision-Oriented-Evaluation-System that you can use for rational allocation of local resources.

The LOCAL DATA is a documentation of LEARNER BENEFITS that either happened or didn't happen as planned.

Such an approach to evaluation will help educators provide effective instruction to all students, including the handicapped. The criteria for this effectiveness will be based upon locally recognized needs and values.

In a nutshell, this research concludes that "you get what you expect when you let co-workers know exactly what you are going to inspect." In other words, let people know the criteria upon which their performance will be judged. You will get two immediate rewards from this prenotification process: first, they will meet more of your criteria; second, they will eventually help you develop better criteria based upon countable learner benefits.